

77 citations found. Retrieving documents...

S. Dumais, J. Platt, D. Heckerman, and M. Sahami. *Inductive learning algorithms and representation for text categorization*. In *Proceedings of CIKM-98, 7th ACM International Conference on Information and Knowledge Management*, pages 148{ 155, 1998.

CiteSeer [Home/Search](#) [Document Not in Database](#) [Summary](#) [Related Articles](#) [Check](#)

This paper is cited in the following contexts:

[First 50 documents](#) [Next 50](#)

[Effective Methods for Improving Naive Bayes Text Classifiers - Kim, Rim, Yook, Lim](#) (Correct)

....nearest neighbor classifiers[7] naive Bayes classifiers[5] and support vector machines[3] etc. **Among these methods, naive Bayes text classifiers have been widely used because of its simplicity although they have been reported as one of poor performing classifiers in text categorization task[8, 2].** Since several studies show that naive Bayes performs surprisingly well in many other domains[1] it is worth of clarifying the reason that naive Bayes fails in the text classification tasks and improving them. **This paper describes the problems in traditional Naive Bayes text classification**

.... $c_j) i=1 \text{ TF}_{ik} P(y_i = c_j | d_i) i=1 \text{ TF}_{is} P(y_i = c_j | d_i) 4)$ The laplacean prior is used to avoid probabilities of zero or one. **For our experiment, parameter is set to 1. This estimation technique has been generally used to implement naive Bayes classifiers in most studies[8, 2, 5].** There are, however, some issues in estimating parameters and calculating scores. **The parameter estimation according to formula (3) regards all of documents belong to c_j as one huge document. In other words, this estimation method does not take into account the fact that there may be important**

[Article contains additional citation **context** not shown here]

S. Dumais, J. Platt, D. Heckerman, and M. Sahami. *Inductive learning algorithms and representation for text categorization*. In *Proceedings of CIKM-98, 7th ACM International Conference on Information and Knowledge Management*, pages 148{ 155, 1998

[Using rough sets to construct sense type decision trees.. - Computing And..](#) (Correct)

....Text categorization, the assignment of natural language texts to one or more predefined categories based on their content, is an important component in information organization and management tasks. **There is an increased interest in developing technologies for automatic text categorization [3].** There are two different ways of approaching the problem: category extraction and category assignment [8] In this paper, we focus on the category extraction problem. **In a previous work [2, 1] following the treatment of strongly typed functional programming languages, we have shown that**

S. Dumais, J. Platt, J. Heckerman, and M. Sahami. *Inductive learning algorithms and representations for text categorization*. In *CIKM'98 - Proc. 7th Intl. Conf. on Information and Knowledge Management*, 1998.

[JHU/APL at TREC 2002: Experiments in Filtering and Arabic.. - Paul McNamee Christine](#) (Correct)

....misspelled words, broken plurals, and infix morphology, and empirically evaluated techniques to overcome them. **Larkey et al. [8] investigated methods for effectively stemming Arabic. apIIIaFah1 0.342 0.104 0.377 0.039 a IIIaFah2 0.342 0.104 0.377 0.039 apII 1Faql 0.059 0.09 0.084 0.369 a IIIaFaql 0.085**

0.118 0.115 0.355 Table 5. **APL Adaptive Results, Assessor topics** Clearly, the heap approach returned too few documents, whereas the queue approach returned too many. **This** is probably mainly due to the much lower amount of feedback. **it** was probably also adversely affected by our choice of

....parameter that guessed too often. **Filtering** Results Discussion in a low training feedback situation, **filtering** seems to require more of a Statistical Language Model score based approach. **Based on the good performance possible in situations with lots of training and feedback (as in TREC 2001) there seems to be a continuum between score based and classification approaches, depending on the amount of training and feedback available.** We conjecture a hybrid approach will be useful to support this continuum. **Given** the successful reports of n gram based retrieval for Arabic, we opted

[Article contains additional citation **context** not shown here]

S. Dumais, J. Platt, D. Heckerman, M. Sahami, *'Inductive Learning Algorithms and Representations for Text Categorization.'* in Proceedings of the 7th International Conference on Information and Knowledge Management (CIKM 98) (1998).

PreBIND and Textomy – mining the biomedical.. - Donaldson.. (2003) (Correct)

....(Fig. 2, item 4) Textomy (Fig. 2, item 7) <http://www.litminer.ca> retrieves these abstracts from PubMed and assigns a score that describes the relative likelihood that the abstract contains molecular interaction information. **Textomy, or text anatomy, is text processing software that uses an SVM [19 21] to capture the statistical pattern of word use in papers that have previously been presented to the machine as papers of interest, in this case, a training set of abstracts that discuss biomolecular interactions.** These SVM scores are stored in the PreBIND database (Fig. 2, item 4) Textomy is

Dumais S, Platt J, Heckerman D and Sahami M *Inductive learning algorithms and representations for text categorization.* Proceedings of the International Conference on Information and Knowledge Management. 1998, 148-155

Comparison of Machine Learning and Traditional.. - Chan, Lee.. (2002) (Correct)

....penalty term regulating the generalization performance of the SVM. **Upon** training, only a fraction of the s will be nonzero. **The** architecture of the SVM in classification is shown in Fig. 1. **SVMs have demonstrated good generalization performance in face recognition [26] text categorization [27], and optical character recognition [28] 29] It has also been applied to data from gene expression [30] DNA and protein analysis [31] 32] D. MOGs** As mentioned in Section I, the generative approach is to model the class conditional density. Since the **input** of the glaucoma data contains

S. T. Dumais, J. Platt, D. Heckerman, and M. Sahami, *"Inductive learning algorithms and representations for text categorization,"* in Proc. ACM-Conf. Information and Knowledge Management (CIKM98), Nov 1998, pp. 148-155.

Integrating Feature and Instance Selection - For Text Classification (Correct)

.... survey on feature and instance selection as two independent problems in the **context** of machine learning is presented in [3] In the **context** of information retrieval and text classification, several works have indicated that effective feature selection can enhance the performance of classifiers, In [5], [11] and [17] a few tens or hundreds of words maximize the performance of a range of classifiers. Similar results are reported in [9] [13] as well. SVMs are a notable exception to this since they

CiteSeer

Find:

Searching for **PHRASE** **search module filtering module lexicon text web pages**

Restrict to: [Header](#) [Title](#) Order by: [Citations](#) [Hubs](#) [Usage](#) [Date](#) Try: [Amazon](#) [B&N](#) [Google \(RI\)](#) [Google \(Web\)](#) [CSB](#) [DBLP](#)

No documents match Boolean query. Trying non-Boolean relevance query.

1000 documents found. Only retrieving 500 documents (System busy - maximum reduced). Retrieving documents... Order: relevance to query.

WebWatcher: Machine Learning and Hypertext - Thorsten Joachims (1995) (Correct) (2 citations)
noted the need for software that helps the user **search** for information. This paper describes the design facing is similar to the problem of Collaborative **Filtering** [Resnick, 1994] The target function we want to is knowledge in the nodes of the graph encoded as **text**. We have begun to explore ways of using this **text**
mobile.csie.ntu.edu.tw/~yjhsu/courses/u1760/papers/webwatcher.ps.gz

An On-Line Cursive Word Recognition System - Seni, Nasrabadi, Srihari (1994) (Correct)
word. Such string is then passed to a procedure **search(ff)** which has knowledge about how to derive ASCII cursive words. The system first uses a **filtering module**, based on simple letter features, to quickly cursive words. The system first uses a **filtering module**, based on simple letter features, to
www.cedar.buffalo.edu/Linguistics/papers/ieee.ps

Results and Challenges in Web Search Evaluation - Hawking, Craswell, Thistlewaite (1999) (Correct)
(11 citations)

Results and Challenges in **Web Search** Evaluation David Hawking
from **pages**. It is not known whether any such **filtering** was applied by the Internet Archive. 2.2 Access is being used in an evaluation framework within the **Text** Retrieval Conference (TREC) and will hopefully
pastime.anu.edu.au/TAR/www8.ps.gz

Human Performance on Clustering Web Pages: A.. - Macskassy, Banerjee.. (1998) (Correct) (5 citations)
using multiple queries or using a topic-specific **search** engine. One way to help in the **search** is by fewer clusters than those with access to the full **text** of each **web page**. Generally the overlap of York, August 1998 1 Human Performance on Clustering **Web Pages: A Preliminary Study** Sofus A. Macskassy,
www.cs.rutgers.edu/~davison/pubs/kdd98.ps

Book Recommending Using Text Categorization with Extracted.. - Mooney (1998) (Correct) (3 citations)
and Building a Database First, an Amazon subject **search** is performed to obtain a list of book-description of computerized matchmaking called collaborative **filtering**. The system maintains a database of the in the AAAI-98/ICML-98 Workshop on Learning for **Text** Categorization and the AAAI-98 Workshop on
ftp.cs.utexas.edu/pub/mooney/papers/libra-textcat98.ps.gz

First-Order Learning for Web Mining - Craven (1998) (Correct) (7 citations)
of learning a concept definition that specifies **search**-control rules for navigating the **Web**. In general, A variety of applications, including information **filtering** systems and browsing assistants, have used context. Cohen [1] has used first-order methods for **text** classification, but the focus was on finding
www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-11/www/wwwkb/ecml98.ps.gz

Personal WebWatcher: design and implementation - Mladenic (1996) (Correct) (2 citations)
to browse the Internet on behalf of a user "It **searches** the World Wide **Web** taking bounded amount of user's shoulder" 4.1 Structure of the learning **module** Learner works in two versions: learning a new Lang [21] developed a system for electronic news **filtering** that uses **text**-learning to generate models of
www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-4/text-learning/www/pww/papers/PWW/pwwTR.ps.Z

Jambalaya: Using Multicast for Blind Distributed Web Searching .. - Navas, Hirsh (1998) (Correct)
Using Multicast for Blind Distributed **Web Searching** and Advertising Julio C. Navas #and Haym
Even when users decide to receive advertisements, **filters** can be used to weed-out unwanted advertisements
o#er can often be just the first few lines of **text** from the document. Unless the author
www.cs.rutgers.edu/pub/technical-reports/dcs-tr-377.ps.Z

Search and Ranking Algorithms for Locating Resources on the.. - Yuwono, Lee (1996) (Correct) (11 citations)
Search and Ranking Algorithms for Locating Resources on

information retrieval, world wide **web** indexing, **text** database 1 Introduction The World Wide **Web** (WWW)
Algorithms for Locating Resources on the World Wide **Web** Budi Yuwono Dik L. Lee Department of Computer and
www.cs.bilkent.edu.tr/~gural/CS550/budidik.ps

Toolkits for a Distributed, Agent-Based Web Commerce System - Guanghao Yan (Correct)
it more time consuming and difficult for people to **search** for information or to locate relevant **web** sites
information selling strategies Communication **module** Transaction processor User interface Products
3-5, 1998. Toolkits for a Distributed, Agent-Based **Web** Commerce System Guanghao Yan Wee-Keong Ng
www.cais.ntu.edu.sg:8000/~wkn/paper/ec98.ps

The MetaCrawler Architecture for Resource Aggregation on the Web - Selberg, Etzioni (1997) (Correct)
(55 citations)
The MetaCrawler Softbot is a parallel **Web** **search** service that has been available at the University
The Harness is implemented as a collection of **modules**, where each **module** represents a particular
users desire, such as phrase **searching** or **filtering** by location, are often absent or require a
www.cs.washington.edu/homes/speed/papers/ieee/ieee-metacrawler.ps

Building a Digital Library for Computer Science Research... - Ian Witten (1996) (Correct) (1 citation)
report archives, and supports a variety of **search** types despite the fact that documents are not
in several respects. First, it provides a full-**text** index of the entire contents of each document,
a large number of documents, many of which are **web** **pages** rather than technical reports. The documents
www.nzdl.org/publications/1996/ACSC.ps

Optimizing complex decision support queries for parallel... - Brunie, Kosch (1997) (Correct)
allocation **module** coupled with a randomized **search** **module** to seek for the best parallelization
It integrates an intelligent resource allocation **module** coupled with a randomized **search** **module** to seek
optimization process. We implemented a first-last **page** cost model, including communication costs. Latency
www.ens-lyon.fr/~hkosch/PDPTA97.ps.Z

Learning to Extract Symbolic Knowledge from the World... - Craven, DiPasquo.. (1998) (Correct) (66 citations)
input URL and explores **pages** using a breadth-first **search** to follow links. Each explored **page** is examined,
Improving learning accuracy in information **filtering**. In International Conference on Machine
to Recognize Class Instances 11 5.1. Statistical **Text** Classification :
www.cs.cmu.edu/~knigam/papers/webkb-tr98.ps.gz

Privacy Interfaces for Information Management - Lau, Etzioni, Weld (1999) (Correct) (7 citations)
her browsing history automatically. A user can **search** her CLIO for **pages** which she has previously
has visited that contain the phrase collaborative **filtering**. To discover pet owners, one might **search**
a match against the document's URL instead of its **textual** content. There is an implicit conjunction over
ftp.cs.washington.edu/tr/1998/02/UW-CSE-98-02-01.PS.Z

Interactive Modular Programming in Scheme - Tung (1992) (Correct) (2 citations)
Abstract This paper presents a **module** system and a programming environment designed to
ftp.cs.indiana.edu/indra/scheme-repository/doc/pubs/imp.ps.gz

Adverbs in the transfer module of MDS - Damova (1995) (Correct)
Adverbs in the transfer **module** of MDS Mariana Damova Universitat Stuttgart
www2.ims.uni-stuttgart.de/~mariana/vm-report-100.ps

Spacetime Constraints Revisited - Ngo, Marks (1993) (Correct) (58 citations)
refine an initial trajectory. We propose a global **search** algorithm that is capable of generating multiple
be described in summary as follows: ffl A dynamics **module** (x2.1) simulates a physically correct virtual
Proceedings, Anaheim, California, August 1993, **pages** 343-350. c fl1993 ACM. Reproduced by permission
www.merl.com/people/marks/spacetime.ps.gz

Machine Learning for Information Extraction in Informal Domains - Freitag (1998) (Correct) (20 citations)
newsgroups where computers are offered for sale in **search** of one that matches a user's specifications. This
which are implemented in Perl, make use of a Perl **module** called Token that defines Perl versions of most
be present in a collection, so that some sort of **filtering** must be performed either before or during
www.cs.cmu.edu/afs/cs/user/dayne/www/ps/diss-freitag.ps.Z

Global Integration of Visual Databases - Wendy Chang (1998) (Correct) (1 citation)

main components include the metadatabase, the **search** agent, and the query manager. The metadatabase the metadata and the templates. Three additional **modules**, metadata collector, template builder and systems are being developed that allow multiple **text** databases to be accessed over the Internet via www.rit.edu/~wcceec/papers/icde98.ps

First 20 documents [Next 20](#)

Try your query at: [Amazon](#) [Barnes & Noble](#) [Google \(RI\)](#) [Google \(Web\)](#) [CSB](#) [DBLP](#)

CiteSeer - citeseer.org - [Terms of Service](#) - [Privacy Policy](#) - Copyright © 1997-2002 [NEC Research Institute](#)

CiteSeer
Research Information Digital Library

Find:

Searching for PHRASE search module micro macro context indexing

Restrict to: [Header](#) [Title](#) Order by: [Citations](#) [Hubs](#) [Usage](#) [Date](#) Try: [Amazon](#) [B&N](#) [Google \(RI\)](#) [Google \(Web\)](#) [CSB](#) [DBLP](#)

No documents match Boolean query. Trying non-Boolean relevance query.

1000 documents found. Only retrieving 250 documents (System busy - maximum reduced). Retrieving documents... Order: relevance to query.

Optimizing complex decision support queries for parallel.. - Brunie, Kosch (1997) (Correct)

allocation module coupled with a randomized search module to seek for the best parallelization
It integrates an intelligent resource allocation module coupled with a randomized search module to seek finding an optimal execution scenario [3] In this context we use randomized search algorithms to present a www.ens-lyon.fr/~hkosch/PDPTA97.ps.Z

Macro and Micro Perspectives of Multistrategy Learning - Reich (1994) (Correct) (6 citations)

nature of the learning method which uses weak search methods directed by the category utility
a learning program for noisy domains that uses one module to extract object descriptions from data and Morgan Kaufmann, San Francisco, CA, 1994 Macro and Micro Perspectives of Multistrategy Learning Yoram or.eng.tau.ac.il:7777/mi4.ps.Z

Sub-element Indexing and Probabilistic Retrieval in the POSTGRES .. - Fontaine (1995) (Correct) (1 citation)

current information retrieval systems use boolean search methods to request and retrieve documents. While words filtered out) are returned to the ranking module which determines the order in which the documents
Sub-element Indexing and Probabilistic Retrieval in the POSTGRES
wuarchive.wustl.edu/packages/postgres/papers/CSD-95-876.ps.Z

Spacetime Constraints Revisited - Ngo, Marks (1993) (Correct) (58 citations)

refine an initial trajectory. We propose a global search algorithm that is capable of generating multiple be described in summary as follows: ffl A dynamics module (x2.1) simulates a physically correct virtual www.merl.com/people/marks/spacetime.ps.gz

Preemption Policy for Hierarchical Cellular Network - Fabrice Valois (1998) (Correct)

served at macrocellular level. The policy then searches a P customer which can be moved from Abstract-In multitier cellular networks, microcells are designed to achieve high spot urban www.prism.uvsq.fr/rapports/1998/document_1998_26.ps.gz

Micro Planning for Mechanical Assembly Operations - Gupta, Paredis, Brown (1998) (Correct)

a tool application position with 1-DOF requires a search through a 1-dimensional search space. To guide Micro Planning for Mechanical Assembly Operations S.
Significant advances have been made in the area of macro planning for assembly operations (i.e.dividing www.cgi.cs.cmu.edu/afs/cs.cmu.edu/user/cjp/www/pubs/ICRA98.ps.gz

An Abstract Interpreter for Improving the Efficiency .. - Ciampolini, Lamma.. (1996) (Correct)

raised. The binding requires to perform a dynamic search in the run-time program representation. For work, we focus on modular logic languages where module composition is performed through union of clauses Apto. 8900, Caracas 1080 A, Venezuela minimal contexts Modular Logic Programs, Static Analysis, www-lia.deis.unibo.it/Research/TechReport/lia96002.ps.Z

Complexity of Finding Alphabet Indexing - Shimozono, Miyano (1995) (Correct)

the problem is NP-complete. Then we give a local search algorithm for this problem and show a result on Complexity of Finding Alphabet Indexing Shinichi Shimozono Department of Control and Q of strings over an alphabet 6, an alphabet indexing /for P Q by an indexing alphabet 0 with j0j ! www.i.kyushu-u.ac.jp/TR/61.ps.Z

Variance and Uncertainty Measures of Population Diversity Dynamics - Mark Bedau (1995) (Correct)

and natural settings, for this facilitates the search for universal features of evolutionary dynamics. Evolving systems have a two-tier structure: a micro level consisting of individuals whose behavior is governed by some explicit dynamics and a macro level consisting of the population as a whole www.sysc.pdx.edu/Faculty/Zwick/Papers/bz1.ps

Scenario Recognition in Airborne Video Imagery - Br mond, Medioni (Correct)

We then illustrate how the scenario recognition **module** works through an example of utilization. Finally, behaviors. This third **module** uses two kinds of **context** (defined as a priori information on the scene iris.usc.edu/Outlines/vsam/wkspBremondMedioni.ps.gz

Duration Calculus of Weakly Monotonic Time - Pandya, Van Hung (1997) (Correct) (11 citations)

of points on the path. Each such point is called a **micro** time point and represented by a tuple (t i) sequence of state changes to occur at a single "**macro**" time point. The resulting notion of time has ftp.iist.unu.edu/pub/techreports/published_papers/paper-report122.ps.gz

Optimization of Three-Dimensional Catalyst Pore Structures - Rieckmann, Keil (1994) (Correct)

optimization calculations of pore structure with a **micro-macro**-pore model [3]In this work ,we have calculations of pore structure with a **micro-macro**-pore model [3]In this work ,we have taken by Eisenstat [2]Optimization The performance **index** was the mole flow of product averaged over the pc50.vt4.tu-harburg.de/pub/doc/veroeff_32.ps.gz

A Multimodal Computer-augmented Interface for Distributed... - Julia, CHEYER (1995) (Correct)

and PDAs)augmented with voice input over a **microphone** or a telephone, depending on where the work with one modality or another. 2. MMAAR (**Micro/Macro** Agent ARchitecture) In this section, we propose ftp.speech.sri.com/pub/people/julia/papers/hcii95.ps.gz

An experimental study of SB-trees - Paolo Ferragina, Roberto Grossi (1996) (Correct)

a practical setting by running a large number of **searching** and updating experiments. We obtain fast K. Genome analysis: Pattern **search** in biological **macromolecules**. In Combinatorial Pattern Matching www.di.unipi.it/~ferragin/Latex/jsoda96.ps.gz

SodaJack: an architecture for agents that search for and... - Geib, Levison, Moore (1994) (Correct) (12 citations)

Science SodaJack: An Architecture For Agents That **Search** For And Manipulate Objects by Christopher Geib Two of these planners are special-purpose **modules** which contribute **context**-specific plans for the are special-purpose **modules** which contribute **context**-specific plans for the tasks of **searching** for ftp.cis.upenn.edu/pub/frcs/tr/94-31.ps.Z

A Bayesian framework for content-based indexing and retrieval - Vasconcelos, Lippman (1998) (Correct) (2 citations)

significant limitations of the query by example **search** paradigm, and all the model parameters can be framework for **indexing** and retrieval in the **context** of large multimedia databases. All the **indexing** A Bayesian framework for content-based **indexing** and retrieval Nuno Vasconcelos and Andrew www.media.mit.edu/~nuno/Papers/BayesRetrieval.ps.gz

acquire-macros: An Algorithm for Automatically Learning... - McGovern (1998) (Correct) (1 citation)

peaks in the evaluation function of a best-first **search** system and constructed **macro**operators out of the and Hierarchy in Reinforcement Learning 1 acquire-macros: An Algorithm for Automatically Learning www.anw.cs.umass.edu/~amy/pubs/mcgovern_nips98_workshop.ps.gz

Reactive Shared Variables Based Systems Syst mes r actifs `a... - Fr d  (Correct)

systems where components (named **modules**) use shared variables to communicate. Each of a global system reaction made out of **module** **micro**-reactions, or alternatively of global instants **modules** appears as first parameter of the System **macro**. **Module** are implemented as RC reactive procedures ftp-sop.inria.fr/meije/rc/rapport19-93.ps

On the Computational Utility of Consciousness - Mathis, Mozer (1995) (Correct) (4 citations)

mapping 1 followed by a slower relaxation **search** (Figure 1)The computational justification for persistent states in a network of computational **modules**. Three simulations are described illustrating here we simply present the framework in the **context** of previous experimental and theoretical work, ftp.cs.colorado.edu/users/mozer/papers/computil.ps

Vector Based Image Matching for Indexing in Case... - Jose, Singh.. (Correct)

abstraction helps the system to avoid an in-depth **search** of those cases which are entirely different from only for those images which contain man-made **macro** objects such as refineries, steel plants etc. The Some of the features of this system in our **context** are the following. 5.1 Support for complex data